The Common Product Distance, A Respective Similarity, and its Application to Curve Fitting

Luciano da Fontoura Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

1st Nov. 2021

Abstract

The common product is a multiset-based signed binary operator (in the mathematical sense of taking two arguments) that provides a quantification of the similarity between two scalars, vectors, functions, or virtually any other mathematical structure. This operation has been shown to impose a more strict quantification of similarity than more commonly employed alternatives including the cosine similarity and cross-correlations. In this work we study a distance derived from the common product, as well as a respective similarity index that have a limited support, therefore reducing or eliminating the influence of outliers. Multiscale, multi-order respective versions of these distance and similarity index are then proposed. To illustrate the potential of the reported concepts and methods, the multiscale similarity index is then applied it to obtain a respective least mean distance approximation method. The obtained method, which is shown to provide several interesting properties specific to certain applications, is then illustrated and discussed.

"*Distant stars, seemingly so similar.*'

LdaFC

1 Introduction

Distance and similarity share many features in science and technology, playing an important role in many concepts, methods, and applications. While the Euclidean distance very probably corresponds to the most frequently adopted distance, the cosine similarity, the inner product and the Jaccard index tend to be frequently employed. The latter index is typically applied to binary or categorical data.

More recently [1, 2, 3], through the consderation of multiset concepts (e.g. [4, 5, 6, 7, 8, 9]), a generalization of the Jaccard index has been obtained that can be applied to real, possibly negative, values. Given that this index has been shown [1] not to take into account the relative interiority of the two compared objects, the index called *coincidence* has been proposed [1, 2, 3] that integrates information from both the real-valued Jaccard and the interiority indices. These two indices, both of which based on the *common product*, have been shown to allow enhanced performance in several tasks, including template matching [10], representation of data as complex networks and highlight of modular structure [11], as well as pattern recognition [12]

The present work focuses on deriving a distance index from the correlates of the common product, real-valued Jaccard, and coincidence indices, which is called coincidence distance. This index quantifies, up to a scale parameter, the distance from a reference value $\bar{c}$ only along an interval $[\bar{c} - w, \bar{c} + w]$. In this manner, points that are further away from x are all considered to have the same distance. This feature provides an interesting approach to limit the influence of outliers on the distance estimation. In other words, the distance quantification operates only locally at a scale defined by the parameter w.

A respective similarity index capable of quantifying the local similarity is then obtained that shares the same ability to consider the similarity only within a window $[x - w, x + w]$ around the reference value x to be compared.

A multiscale curve fitting method is then derived from this similarity index that has the following interesting properties: (i) it operates locally, ignoring points that are further away than w, which are understood as outliers to be left out; (ii) the definition of the window to be considered can be readily controlled by the scaling parameter w; (iii) the sharpness of the active region can be conveniently controlled by a parameter D; (iv) provided the original points are not too noisy, the fitting solution can be effectively obtained by using the simple gradient

descent approach, otherwise local peaks may occur and simulated annealing approaches, or the Hough transform (e.g. [13, 14]). can be considered.

2 The Common Product Distance

The *common product* between two scalars x and y has been defined [1, 2, 3, 15] as:

$$x \sqcap y = s_{xy} \min\{s_x x, s_y y\} \tag{1}$$

with $x \sqcap y = y \sqcap x$.

This product has been derived from the extension of multiset theory to real values [1, 2, 3, 15] and shown to be directly related to the Kronecker delta function [3], though providing a more practical quantification of the similarity between two scalars.

A better understanding of this product can be obtained by fixing one of its arguments to a constant value. For instance, in case the argument y in Equation 6 is fixed as $y = \bar{c} \neq 0$, the binary operation becomes a function of x.

Figure 1 illustrates the common product for $x = 2$, i.e. $2 \sqcap y = y \sqcap 2$.

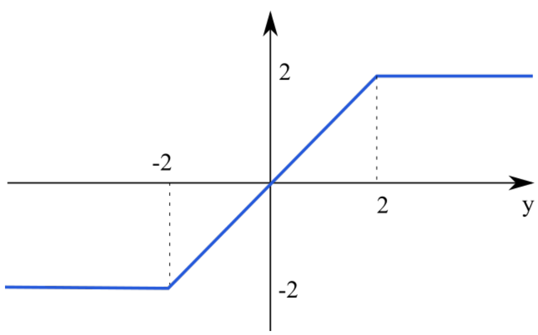

Figure 1: The common product binary operator with one of its arguments fixed, i.e. $x = 2$, therefore implementing the single variable function $2 \sqcap y = y \sqcap 2$ shown in this figure.

Observe that $-\bar{c} \leq x \sqcap \bar{c} \leq \bar{c}$. A normalized version of the common product can be obtained as:

$$\bar{c}\bar{\sqcap}y = \frac{s_{\bar{c}y} \min\{s_{\bar{c}}\bar{c}, s_y y\}}{\bar{c}} \tag{2}$$

with $-1 \leq x\bar{\sqcap}y \leq 1$ and $\bar{c} \neq 0$.

This version of the common product is shown in Figure 2(a).

It is also possible to take the absolute value of the above similarity index, yielding:

$$\bar{c}\hat{\sqcap}y = \left| \frac{\min\{s_{\bar{c}}\bar{c}, s_y y\}}{\bar{c}} \right| \tag{3}$$

and now we have with $0 \leq x\hat{\sqcap}y \leq 1$. This function is illustrated in Figure 2(b).

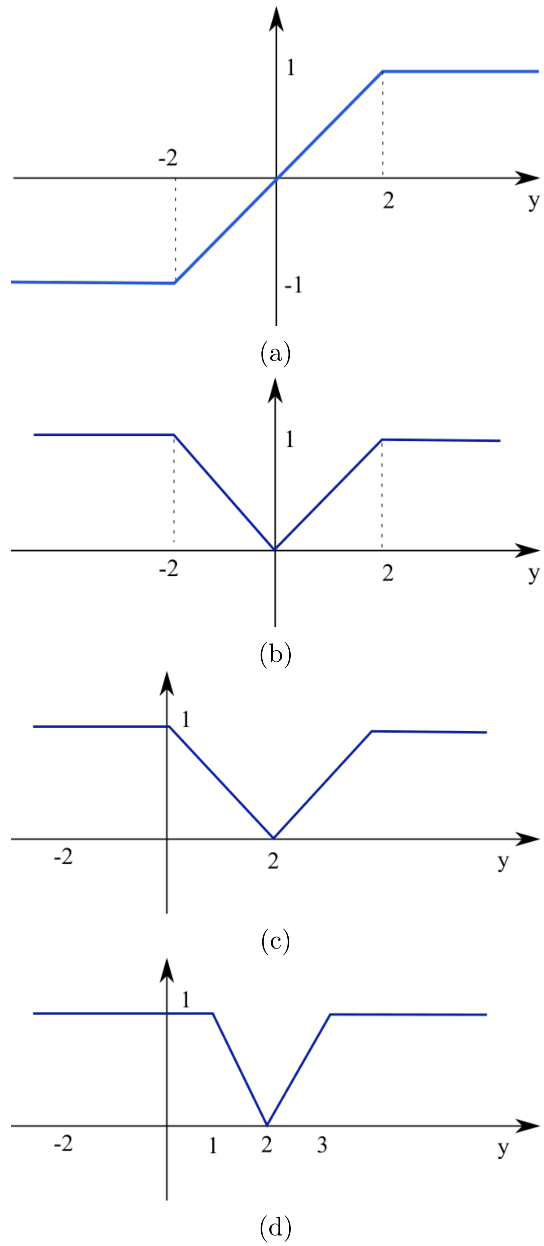

Figure 2: Successive transformations of the common product with one argument kept constant at $y = 2$ in Fig. 1 (a) so as to have unit height (a), absolute value (b), displaced to $x = 2$ (c), width 2 (d).

A shifted version of the previous function can now be derived as:

$$\hat{\Delta}(\bar{c}, y) = \left| \frac{\min\{s_{\bar{c}}\bar{c}, s_{y-\bar{c}}(y - \bar{c})\}}{\bar{c}} \right| \tag{4}$$

with $0 \leq \bar{c}\hat{\sqcap}y \leq 1$. We also have that $\hat{\Delta}(x, \bar{c}) = \hat{\Delta}(\bar{c}, x)$. This function is shown in Figure 2(c).

In order to allow the width of the above distance to be controlled by a parameters w, implying width $2w$, we rewrite the previous equation as;

$$\Delta(x, \bar{c}, w) = \left| \frac{\min\left\{s_{\bar{c}}\bar{c}, s_{(y-\bar{c})\bar{c}/w}((y - \bar{c})\bar{c}/w)\right\}}{\bar{c}} \right| \tag{5}$$

This distance is illustrated in Figure 2(d).

The function $\delta(x, \bar{c})$ can be brought back to its original general binary form as:

$$\Delta(x, y, w) = \left| \frac{\min\left\{ s_{\bar{x}}\bar{x}, s_{(y-\bar{x})\bar{x}/w}((y-\bar{x})\bar{x}/w) \right\}}{\bar{x}} \right| \quad (6)$$

with $0 \leq \Delta(x, y, w) \leq 1$, $y \neq 0$ and $x \neq 0$.

Therefore, we have that the above distance has been obtained all the way from the common product through successive transformations. For this reason, in this work this distance is called the *common product distance.*

The obtained distance has some interesting specific properties that makes it an interesting option for certain applications. First, it is particularly simple, involving just a minimum and a division operation, plus the very simple sign functions. Of particular importance, though, is the saturation of the distance for values of y that are larger than w. The latter feature implies all the more distant points to be considered all with the maximum distance of 1, therefore reducing the influence of outliers. As such, the obtained distance can provide an interesting alternative in situations in which the effect of outliers need to be reduced. Observe also that the obtained distance is multiscale, in the sense that its width is controlled by the parameter w.

Given the interesting features obtained for $\Delta(x, y, w)$, a respectively derived similarity index can be considered. This can be immediately obtained by making:

$$s(x, y, w) = 1 - \Delta(x, y, w) \quad (7)$$

As with the common product distance, this similarity index is also multiscale and can cope with outliers if needed. Indeed, the similarity with outliers, i.e. points that are further away than w becomes zero, consequently not being taken into account. This feature allows the introduction of a new parameter D controlling the order (or degree) of the similarity index, i.e.:

$$s(x, y, w, D) = [1 - \Delta(x, y, w)]^D \quad (8)$$

with $0 \leq s(x, y, w, D) \leq 1$.

Though the parameter D can take any real value, here we limit our attention to positive integer values of D.

Figure 3 depicts the common product similarity for $\bar{c}$, $w = 4$ and $d = 1, 2, \ldots, 8$.

Observe the effect of the parameter D on the sharpness of the similarity peak . Therefore, more strict similarity measurements are obtained by using larger values of D.

3 Application to Line Fitting

In this section we provide an illustration of the potential of the developed common product similarity with respect to the important task of straight line fitting.

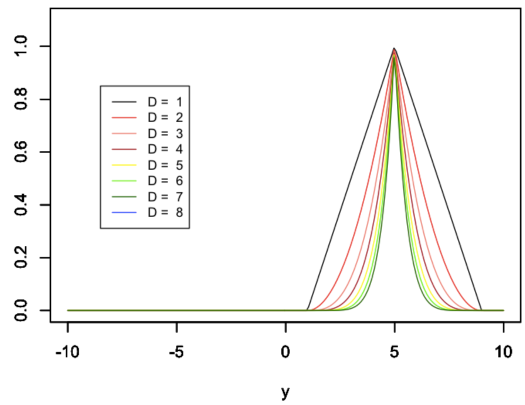

Figure 3: The multiscale *common product similarity* for $\bar{c} = 5$, $w = 4$ and $D = 1, 2, \ldots, 8$.

An affine, or straight function, is henceforth understood to correspond to:

$$y = m\ x + c \quad (9)$$

where the parameter m is commonly called the line *slope* and the parameter c its respective *intersect.*

The polar line equation can also be of interest, being expressed as:

$$\rho = x\cos(\theta) + y\sin(\theta) \quad (10)$$

Fitting a line to a set of discrete points typically starts with a table containing the respective x and y coordinates of the latter. The traditional approach to line fitting has been through the least square methods (e.g. [16]), though the respective L_1 norm (absolute value) version — least absolute deviations – LAV (e.g. [17]) — has also been applied (e.g. [16]). However, neither of these can cope with outliers or have immediate parameters controlling the function profile.

In the present work, we apply the common product similarity developed from the multiset-based common product in Section 2

The basic idea is to apply the common product similarity to gauge the similarity between each candidate solution

$$y_c = m_c x_o + c_c \quad (11)$$

and the y_o coordinate of each of the N original points (x_o, y_o), i.e. $s(y_c, y_o, w, D)$.

The optimization criterion therefore consists of maximizing the sum of the respective common product similarities, i.e.:

Find m_c and c_c that maximizes:

$$\sum_{i=1}^{N} s(y_{c,i}, y_{o,i}, w, D) \quad \textit{for } i = 1, 2, \ldots, N \quad (12)$$

The especially important issue here regards the method to be applied for implementing the above optimization. There are several possibilities, but gradient ascent has been experimentally found to be particularly effective given the typically smooth energy landscapes that are obtained for relatively large values of w.

Figure 4 illustrates a set of points to be fitted by a straight line.

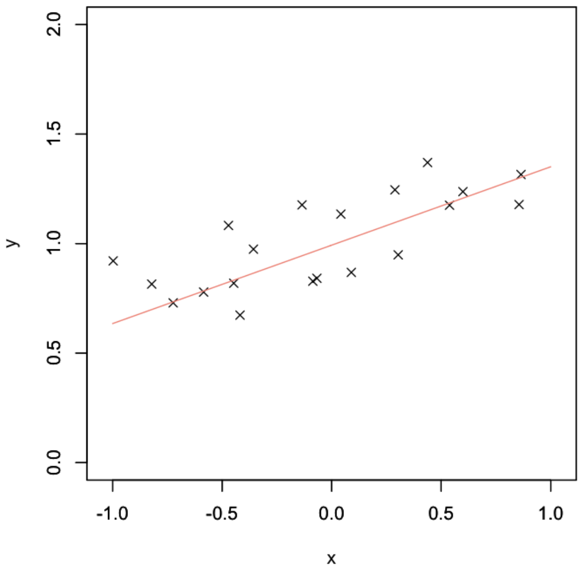

Figure 4: A set of points to be fitted by a straight line $y = mx + c$. Also shown in salmon is the result of the fit by using $w = 0.5$ and $D = 2$. These points were obtained by adding symmetric uniform noise to a straigth line defined by $m_o = 0.3$ and $c_o = 1$.

The distribution of the similarity values along the parameters space $[m, c]$ can be effectively appreciated when displayed as an image. Figure 5 illustrates the accumulated similarity along the parameter space obtained by considering $w = 1$ and $D = 2$ in the set of points in Fig. 4. For simplicity's sake, only the portion of the parameter space respective to $m \in [-1, 1]$ is henceforth shown.

Despite the relatively high level of noise in the original points, a remarkably smooth surface has been obtained. The surface of accumulated similarities present a peak that coincides accurately with the original line parameters $m_o = 0.3$ and $c_o = 1$ Such a smooth surface with such a well defined peak tends to substantially favor optimization methods such as gradient ascent.

Figure 6 illustrates the parameter space obtained for $w = 3$ and a substantially higher $D = 10$. The obtained similarity surface is still quite smooth, though with a sharper peak implied by $D = 10$. This type of surface would still be suitable for gradient ascent approaches.

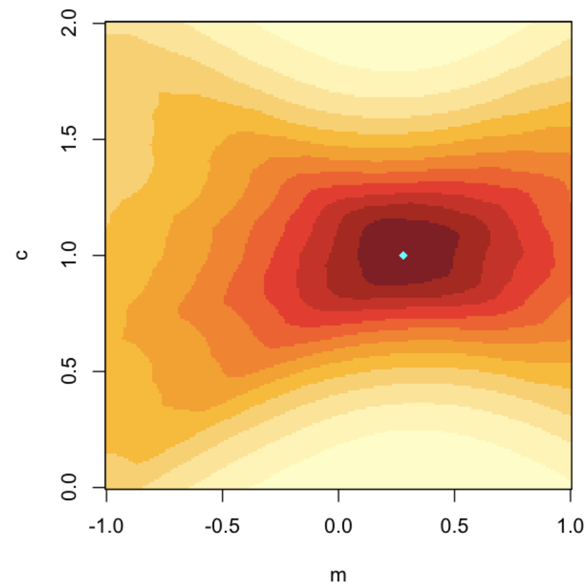

Figure 5: The parameters space $[m, c]$ obtained for the set of points in Fig. 4 with respect to $w = 1$ and $D = 2$. A remarkably smooth surface has been obtained even considering the relatively high level of noise in the original points, which substantially favors optimization approaches such as gradient ascent. It is observed that the obtained surface contains much more levels than shown in this figure, for the level sets have been imposed in order to give a better perspective of the geometry of the surface. The point in cyan indicates the position of the optimal, original parameters $m_o = 0.3$ and $c_o = 1$.

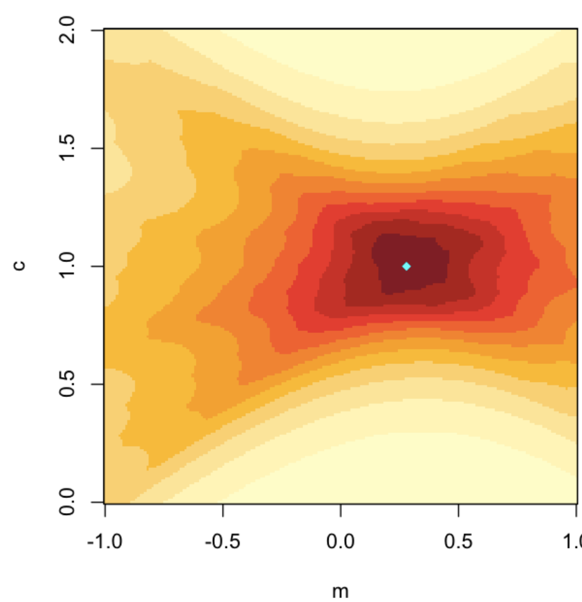

Figure 6: The parameters space $[m, c]$ obtained for the set of points in Fig. 4 with respect to $w = 3$ and $D = 10$. A still remarkably smooth surface has been obtained, though with a sharper peak than in the previous case. This example illustrates the effect of the parameter D in controlling how much strict the similarity quantification is being performed.

To complete our examples, we present in Figure 7 the similarity surface on the parameter space obtained for $w = 0.3$ and $D = 1$, which implies a much narrower window than in the two above examples. Remarkably, the obtained similarity surface still resulted, at least for the considered set of points, intensely smooth, with an even sharper peak around the optimal original parameters shown in cyan.

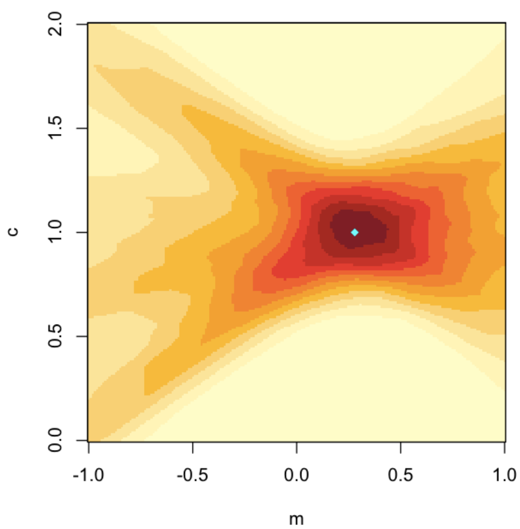

Figure 7: The parameters space $[m, c]$ obtained for the set of points in Fig. 4 with respect to $w = 0.3$ and $D = 1$. A still remarkably smooth surface has been obtained, though with an even sharper peak than in the previous cases.

Figure 8 illustrates the application of the gradient ascent for $w = 1$ and $D = 2$. The trajectory of the gradient, shown in magenta, indicates a relatively direct convergence to the peak. As a consequence of the loss of information implied by the high level of noise added to the original points, the found peak is not perfectly identical to the original noiseless line parameters $m_o = 0.3$ and $c_o = 1$.

Figure 9 depicts the resul obtained for $w = 0.3$ and $D = 1$. Again, the gradient ascent was relatively direct, reaching a peak value very similar to that obtained in the previous gradient ascent example.

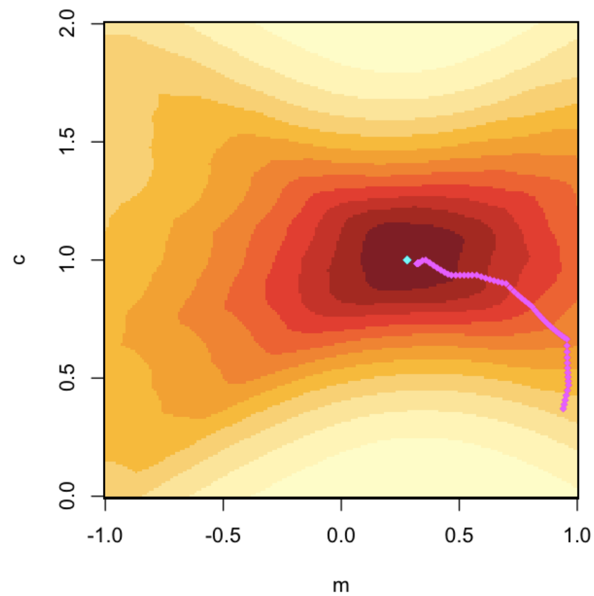

Figure 8: The trajectory of the gradient ascent obtained for $w = 1$ and $D = 2$ respectively to the points in Fig. 4. In addition to the relatively direct trajectory towards the peak, the obtained values, $m = 0.347$ and $c = 0.985$, are very close to the parameters of the original noiseless image from which the points were obtained through addition of elevated level of noise.

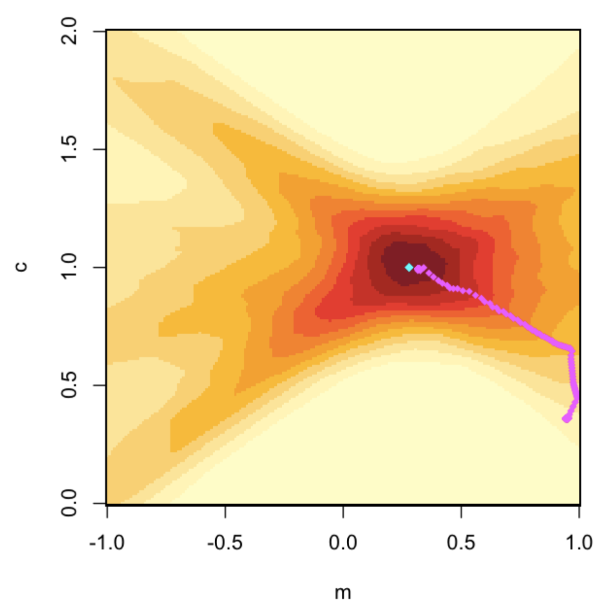

Figure 9: The trajectory of the gradient ascent obtained for $w = 1$ and $D = 2$ respectively to the points in Fig. 4. In addition to the relatively direct trajectory towards the peak, the obtained values, $m = 0.34$ and $c = 0.994$, are very close to the parameters of the original noiseless line.

4 Concluding Remarks

Developed by using multiset concepts, the common product has verified to allow enhanced performance in a large number of situations, mainly as a consequence of implementing a more strict quantification of similarity than the extensively used cosine similarity, inner product, and cross-correlation.

In the present work, we started with the common product and developed, through successive modifications, a respective distance, here called the common product dis-

tance, which has several interesting features including multiscale operation (controlled by the parameter w), adjustable sharpness (through parameter D), ability to reduce the impact of outlier points, as well as conceptual and arithmetic relative simplicity.

It was then shown that this distance could be brought back to the similarity perspective with some interesting features, such as operating completely in the local scale defined by the width $2w$. This feature is of particular interest in case the outliers are to be completely avoided, as it may happen when performing line fitting.

The potential of the proposed concepts was then illustrated with respect to the important problem of line fitting, with interesting and promising results that makes this method interesting in certain applications, such as when dealing with high levels of noise or avoiding outliers.

The proposed curve fitting methodology has not been compared to the traditional least mean square approach, or even the absolute value related method, because each of these approaches are here understood to serve specific demands in the light of their intrinsic properties.

Further developments include a more systematic evaluation of the method with respect to varying levels and types of noise, as well as the effect of the involved parameters w and D. It would also be interesting to develop further the possibility of using the Hough transform as a means of finding the best fit, and compare this with the gradient descent approach. The problem of local peaks in the parameter space also deserved further investigation

Acknowledgments.

Luciano da F. Costa thanks CNPq (grant no. 307085/2018-0) and FAPESP (grant 15/22308-2).

References

[1] L. da F. Costa. Further generalizations of the Jaccard index. https://www.researchgate.net/publication/355381945_Further_Generalizations_of_the_Jaccard_Index, 2021. [Online; accessed 21-Aug-2021].

[2] L. da F. Costa. Multisets. https://www.researchgate.net/publication/355437006_Multisets, 2021. [Online; accessed 21-Aug-2021].

[3] L. da F. Costa. On similarity. https://www.researchgate.net/publication/355792673_On_Similarity, 2021. [Online; accessed 21-Aug-2021].

[4] J. Hein. *Discrete Mathematics.* Jones & Bartlett Pub., 2003.

[5] D. E. Knuth. *The Art of Computing.* Addison Wesley, 1998.

[6] W. D. Blizard. Multiset theory. *Notre Dame Journal of Formal Logic*, 30:36—66, 1989.

[7] W. D. Blizard. The development of multiset theory. *Modern Logic*, 4:319–352, 1991.

[8] P. M. Mahalakshmi and P. Thangavelu. Properties of multisets. *International Journal of Innovative Technology and Exploring Engineering*, 8:1–4, 2019.

[9] D. Singh, M. Ibrahim, T. Yohana, and J. N. Singh. Complementation in multiset theory. *International Mathematical Forum*, 38:1877–1884, 2011.

[10] L. da F. Costa. Comparing cross correlation-based similarities. https://www.researchgate.net/publication/355546016_Comparing_Cross_Correlation-Based_Similarities, 2021. [Online; accessed 21-Oct-2021].

[11] L. da F. Costa. Coincidence complex networks. https://www.researchgate.net/publication/355859189_Coincidence_Complex_Networks, 2021. [Online; accessed 21-Aug-2021].

[12] L. da F. Costa. Real-valued jaccard and coincidence based hierarchical clustering. https://www.researchgate.net/publication/355820021_Real-Valued_Jaccard_and_Coincidence_Based_Hierarchical_Clustering, 2021. [Online; accessed 21-Aug-2021].

[13] P.V.C. Hough. Method and means for recognizing complex patterns, December 18 1962. US Patent 3,069,654.

[14] L. da F. Costa. *Shape Classification and Analysis: Theory and Practice.* CRC Press, Boca Raton, 2nd edition, 2009.

[15] L. da F. Costa. Generalized probabilities. https://www.researchgate.net/publication/356075428_Generalized_Probabilities, 2021. [Online; accessed 21-Aug-2021].

[16] W.H Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, 3rd edition, 2007.

[17] I. Barrodale and F. D. K. Roberts. An improved algorithm for discrete l1 linear approximation. *SIAM Journal on Numerical Analysis*, 10(5):839–848, 1973.